CLAIMS:

1.   A Link Packet Scheduler for each virtual lane (VL) at a given port, comprising:

a N-bit counter arranged to accumulate free credits relinquished, when a data packet is

removed from a receive buffer and that buffer space is reclaimed as available for data packet

storage, or when a link packet is received whose Flow Control Total Bytes Sent (FCTBS) field

differs from actual blocks received (ABR) at the given port;

a first comparator arranged to make comparison between accumulated free credits from

the N-bit counter and a programmable credit threshold;

a second comparator arranged to make comparison between a current buffer receive

utilization indicating a data storage level of the receive buffer and a programmable utilization

threshold; and

a logic device arranged to track a current link state of the corresponding port, to monitor

the amount of receive buffer resources from the first and second comparators and to schedule for

the transmission of a link packet, via a physical link.

2.   The Link Packet Scheduler as claimed in claim 1, wherein said logic device

comprises:

a first OR gate arranged to track whether a Link State Machine transitions into one of the

LinkInitialize, the *LinkArm* and the *LinkActive* state, whether a configuration strap for enabling

1   *loopback* operation changes states, and whether a per-VL Link Packet Watchdog Timer expires,

2   and generate therefrom a logic signal;

3   an AND gate arranged to logically combine outputs of the first comparator and the

4   second comparator and produce a logic signal; and

5   a second OR gate arranged to logically combine the logic signals from the first OR gate

6   and the AND gate and generate an indication for transmission of the link packet for the virtual

7   lane (VL) on the given port.

1   3.   The Link Packet Scheduler as claimed in claim 2, wherein said Link Packet

2   Watchdog Timer is utilized to ensure that at minimum link packets will be independently

3   scheduled at least once every 65,536 symbol times in accordance with the InfiniBand™

4   specification.

1   4.   The Link Packet Scheduler as claimed in claim 1, wherein each of said credits is

2   defined to be 64-bytes of available receive buffer space.

1   5.   The Link Packet Scheduler as claimed in claim 1, wherein said link packet

2   contains at least a Flow Control Total Block Sent (FCTBS) field which is generated by the

3   transmitter logic and is set to the total number of blocks (64 bytes) transmitted since link

4   initialization, and a Flow Control Credit Limit (FCCL) field which is generated by the receiver

5   logic and is used to grant transmission credits to the remote transmitter.

1   6.   The Link Packet Scheduler as claimed in claim 5, wherein said Flow Control

2   Total Block Sent (FCTBS) field and said Flow Control Credit Limit (FCCL) field are used to

3   guarantee that data is never lost due to lack of receive buffer at each end of a physical link.

1   7.   The Link Packet Scheduler as claimed in claim 1, wherein said programmable

2   utilization threshold is set such that the receive buffer has multiple data packets pending

3   processing.

1   8.   A data network, comprising:

2   a host system having a host-fabric adapter;

3   at least one remote system;

4   a switch fabric which interconnects said host system via said host-fabric adapter to said

5   remote system along different physical links for data communications; and

6   one or more communication ports provided in said host-fabric adapter of said host system

7   each port including a set of transmit and receive buffers capable of sending and receiving data

8   packets concurrently via respective transmitter and receiver at an end of a physical link, via said

9   switched fabric, and a flow control mechanism utilized to prevent loss of data due to receive

10   buffer overflow at the end of said physical link.

1   9.   The data network as claimed in claim 8, wherein said flow control mechanism is

2   configured to support flow control through multiple virtual lanes VLs on the given port and to

3   perform:

4   determining when a Link State Machine transitions into one of a *LinkInitialize* state, a

5   *LinkArm* state and a *LinkActive* state, or when a configuration strap for enabling *loopback*

6   operation changes states;

7   if the Link State Machine transitions into one of the *LinkInitialize* state, the *LinkArm*

8   state and a *LinkActive* state, or when the configuration strap for enabling *loopback* operation

9   changes states, scheduling a link packet transmission for all supported virtual lanes VLs on a

10   given port;

11   if the Link State Machine does not transition into one of the *LinkInitialize* state, the

12   *LinkArm* state and a *LinkActive* state, or when the configuration strap for enabling *loopback*

13   operation does not change states, determining whether a Link Packet Watchdog Timer per

14   virtual lane VL has expired;

15   if the per VL Link Packet Watchdog Timer has expired, scheduling a link packet

16   transmission for that virtual lane VL;

17   if the per VL Link Packet Watchdog Timer has not expired, determining whether a

18   receive buffer utilization indicating a data storage level of the receive buffer exceeds a

19   programmable utilization threshold;

20   if the receive buffer utilization exceeds the programmable utilization threshold,

21   prohibiting the link packet transmission for that virtual lane VL;

37

1   if the receive buffer utilization does not exceed the programmable utilization threshold,

2   determining whether free credits accumulated exceeds a programmable credit threshold;

3   if the free credits accumulated exceeds the programmable credit threshold, scheduling a

4   link packet transmission for that virtual lane VL; and

5   if the free credits accumulated does not exceed the programmable credit threshold,

6   prohibiting the link packet transmission for that virtual lane VL.

1   10.   The data network as claimed in claim 9, wherein each of said credits is 64-bytes

2   from the receive buffer, and said credits are relinquished when data packets are removed from

3   the receive buffer and that space is reclaimed as available for packet storage, or when link

4   packets are received whose Flow Control Total Bytes Sent (FCTBS) fields differ from actual

5   blocks received (ABR) at the given port.

1   11.   The data network as claimed in claim 8, wherein said flow control mechanism

2   contains a Link Packet Scheduler per virtual lane (VL) arranged to schedule a link packet

3   transmission for that virtual lane VL.

1   12.   The data network as claimed in claim 11, wherein said Link Packet Scheduler

2   comprises:

3   a N-bit counter arranged to accumulate free credits relinquished, when a data packet is

4   removed from a receive buffer and that buffer space is reclaimed as available for data packet

1   storage, or when a link packet is received whose Flow Control Total Bytes Sent (FCTBS) field

2   differs from actual blocks received (ABR) at the given port;

3   a first comparator arranged to make comparison between accumulated free credits from

4   the N-bit counter and a programmable credit threshold;

5   a second comparator arranged to make comparison between a current buffer receive

6   utilization indicating a data storage level of the receive buffer and a programmable utilization

7   threshold; and

8   a logic device arranged to track the current link state of the corresponding port, to

9   monitor the amount of receive buffer resources from the first and second comparators and to

10   schedule the transmission of a link packet, via a physical link.

1   13.   The data network as claimed in claim 12, wherein said logic device comprises:

2   a first OR gate arranged to track whether a Link State Machine transitions into one of the

3   *LinkInitialize*, the *LinkArm* and the *LinkActive* state, whether a configuration strap for enabling

4   *loopback* operation changes states, and whether a per-VL Link Packet Watchdog Timer expires

5   at a predetermined symbol time, and to produce therefrom a logic signal;

6   an AND gate arranged to logically combine outputs of the first comparator and the

7   second comparator and to produce a logic signal; and

8   a second OR gate arranged to logically combine the logic signals from the first OR gate

9   and the AND gate and to produce an indication for transmission of the link packet for the virtual

10   lane (VL) on the given port.

1   14.   The data network as claimed in claim 12, wherein said Link Packet Watchdog

2   Timer is utilized to ensure that at minimum link packets will be independently scheduled at least

3   once every 65,536 symbol times in accordance with the InfiniBand™ specification.

1   15.   The data network as claimed in claim 12, wherein each of said credits is defined

2   to be 64-bytes of available receive buffer space.

1   16.   The data network as claimed in claim 12, wherein said link packet contains at

2   least a Flow Control Total Block Sent (FCTBS) field which is generated by the transmitter logic

3   and is set to the total number of blocks (64 bytes) transmitted since link initialization, and a Flow

4   Control Credit Limit (FCCL) field which is generated by the receiver logic and is used to grant

5   transmission credits to the remote transmitter.

1   17.   The data network as claimed in claim 16, wherein said Flow Control Total Block

2   Sent (FCTBS) field and said Flow Control Credit Limit (FCCL) field are used to guarantee that

3   data is never lost due to lack of receive buffer at each end of a physical link.

1   18.   The data network as claimed in claim 12, wherein said programmable utilization

2   threshold is set such that the receive buffer has multiple data packets pending processing.

1   19.   A method of flow control of a link packet in a host-fabric adapter installed in a

2   data network, comprising:

3   determining when a Link State Machine transitions into one of a *LinkInitialize* state, a

4   *LinkArm* state and a *LinkActive* state, or when a configuration strap for enabling *loopback*

5   operation changes state;

6   if the Link State Machine transitions into one of the *LinkInitialize* state, the *LinkArm*

7   state and a *LinkActive* state, or when the configuration strap for enabling *loopback* operation

8   changes states, scheduling transmission of a link packet for all supported virtual lanes VLs on a

9   given port;

10   if the Link State Machine does not transition into one of the *LinkInitialize* state, the

11   *LinkArm* state and a *LinkActive* state, or when the configuration strap for enabling *loopback*

12   operation does not change states, determining whether a Link Packet Watchdog Timer per

13   virtual lane VL has expired;

14   if the per VL Link Packet Watchdog Timer has expired, scheduling transmission of a link

15   packet for that virtual lane VL;

16   if the per VL Link Packet Watchdog Timer has not expired, determining whether a

17   receive buffer utilization indicating a data storage level of the receive buffer exceeds a

18   programmable utilization threshold;

19   if the receive buffer utilization exceeds the programmable utilization threshold,

20   prohibiting transmission of a link packet for that virtual lane VL;

1   if the receive buffer utilization does not exceed the programmable utilization threshold,

2   determining whether free credits accumulated exceeds a programmable credit threshold;

3   if the free credits accumulated exceeds the programmable credit threshold, scheduling

4   transmission of a link packet for that virtual lane VL; and

5   if the free credits accumulated does not exceed the programmable credit threshold,

6   prohibiting transmission of a link packet for that virtual lane VL.

1   20.   The method as claimed in claim 19, wherein each of said credits is 64-bytes from

2   the receive buffer, and said credits are relinquished when data packets are removed from the

3   receive buffer and that space is reclaimed as available for packet storage, or when link packets

4   are received whose Flow Control Total Bytes Sent (FCTBS) fields differ from actual blocks

5   received (ABR) at the given port.

1   21.   The method as claimed in claim 19, wherein said Link Packet Watchdog Timer is

2   utilized to ensure that at minimum link packets will be independently scheduled at least once

3   every 65,536 symbol times in accordance with the InfiniBand™ specification.

1   22.   The method as claimed in claim 19, wherein each of said credits is defined to be

2   64-bytes of available receive buffer space.

1   23.   The method as claimed in claim 19, wherein said link packet contains at least a

2   Flow Control Total Block Sent (FCTBS) field which is generated by the transmitter logic and is

3   set to the total number of blocks (64 bytes) transmitted since link initialization, and a Flow

4   Control Credit Limit (FCCL) field which is generated by the receiver logic and is used to grant

5   transmission credits to the remote transmitter.

1   24.   The method as claimed in claim 19, wherein said Flow Control Total Block Sent

2   (FCTBS) field and said Flow Control Credit Limit (FCCL) field are used to guarantee that data

3   is never lost due to lack of receive buffer at each end of a physical link.

1   25.   The method as claimed in claim 19, wherein said programmable utilization

2   threshold is set such that the receive buffer has multiple data packets pending processing.